



Washington OSPI WIDA Screener Identification Criteria Setting

July 2021

Susan Davis-Becker, Ph.D. &
Chad Buckendahl, Ph.D.
sdavisbecker@acsventures.com
cbuckendahl@acsventures.com
402.770.0085

Overview

The Washington Office of Superintendent of Public Instruction (OSPI) contracted with ACS Ventures, LLC (ACS) to establish criteria for interpreting scores for the WIDA Screener Online Assessment. ACS worked with OSPI leadership to design and facilitate a study by which Washington subject matter experts (educators) could make judgments as to what level(s) of performance would indicate that a student is identified and eligible for English Language (EL) services. This report documents the process and results of these efforts.

The WIDA Screener Online is an assessment tool designed to determine if students should be eligible for EL services. The assessment includes measures of four domains: Reading, Listening, Writing, and Speaking. The Reading and Listening domains include multiple-choice items (scored by the computer) whereas the Writing and Speaking domains include performance tasks (scored by the test administrator). Tasks and items are selected for a student based on their prior performance (i.e., students who do well receive more challenging tasks).

Students receive a score report from the WIDA Screener Online with a total of seven scores (three composite, four domain):

- Composite Scores
 - Overall: Reading (35%) + Writing (35%) + Speaking (15%) + Listening (15%)
 - Oral Language: Speaking (50%) + Listening (50%)
 - Literacy: Reading (50%) + Writing (50%)
- Domain Scores
 - Reading
 - Writing
 - Speaking
 - Listening

Each score is reported on the WIDA performance scale (1-6, whole integers for domain scores, half points possible for composite scores):

- 1) Entering
- 2) Emerging
- 3) Developing
- 4) Expanding
- 5) Bridging
- 6) Reaching

The expectations for each level are described generally in the *Performance Level Definitions* and more specifically for each grade level in the *Can-Do Descriptors* provided by WIDA.

Each state that uses the WIDA Screener Online is responsible for establishing criteria by which students can be identified as English Learners. To complete this activity, ACS and OSPI had access to the general and more specific performance expectations at each level, some high-level student data from other states, and the identification criteria used by other states. Therefore, ACS implemented a judgmental process by which Washington subject matter experts could review the performance expectations and determine what level of performance signaled that a student did not need English language services.



Meeting Process

OSPI selected eighteen subject matter experts to participate in this study who all worked with English language learners in Washington. Based on their experience, the panelists were organized into six panels to reflect the different grade levels of the WIDA Screener: K, 1, 2-3, 4-5, 6-8, 9-12. Table 1 below shows the experience of each panel.

Table 1. Subject Matter Expert Background

Panel	Count	Average Years of Experience in Education	Average Years of Experience working with EL students
K	3	10.00	5.67
1	3	13.67	13.00
2-3	3	22.33	18.33
4-5	3	27.50	23.50
6-8	3	13.33	13.33
9-12	3	25.67	23.00

To complete this task, ACS designed and facilitated a series of meetings July 6-9, 2021, where the subject matter experts had the opportunity to discuss general expectations as well as more specific grade level criteria while iterating on their judgments. The process followed is outlined below by meeting day.

1. Pre-meeting orientation – July 6 (2 hours)

During this initial brief meeting, panelists were asked to test their connection to the virtual meeting platform (Zoom) and confirm they could access the online survey tool that would be used for gathering opinions throughout the meeting (Microsoft Forms) by completing a demographic form and nondisclosure agreement. OSPI provided a welcome to the panel and ACS provided a brief introduction as to the purpose of the study including the purpose of the WIDA screener, the need to identify the eligibility criteria, and the types of judgments the panel will be asked to make during the study.

2. Meeting Day 1: Process Training and General Discussion – July 7 (4 hours)

The first full meeting day was conducted as a general session with all panelists. ACS conducted training for the entire process that would be followed (goals and activities for each meeting day), the WIDA Screener Online (structures, items and task types, scoring), and how they would make their judgments that would eventually lead to the panel's final recommendations. ACS facilitated three brainstorming activities. Specifically, the panel discussed:

- what signals they use to determine that a student does not need English language services.
- what evidence within the WIDA *performance level definitions* (PLDs) matched the signals they discussed in the first activity.
- which scores from the WIDA Screener Online should be used as part of the identification criteria. The panelists made an initial independent judgment on this specific question, discussed the results, and then made their final judgments.

3. Meeting Day 2: Grade Level Group Discussions – July 8 (4 hours)

During this meeting, ACS presented a training to all panelists on the process and goals for the day along with the results of the discussions and voting from the previous meeting day. Panelists then spent the remainder of the meeting time with their grade level group (K, 1, 2-3, 4-5, 6-8, 9-12) completing the following activities:



- Review the *Can-Do* descriptors from their grade level and review which expectations align with the general signals discussed on day 1.
- Make and discuss their initial judgments as to what levels of performance signals that a student does not need EL services. Evaluate their judgments against summary data from other states that includes the average domain scores for students who achieved a given overall score. This allowed each group to evaluate their preliminary recommendations against real data from the WIDA Screener assessments.
- Review the perspectives shared from other grade-level groups and finalize their expectations.

4. Meeting Day 3: Cross Grade Articulation

For the final day of the meeting, one panelist from each grade level group was invited to participate and share the perspectives of their group to have the panels come to consensus on their recommendations. Each representative presented their group’s perspectives and ACS facilitated a discussion by highlighting the common and differing perspectives.

Meeting Results

The panel was tasked with making two types of recommendations as a part of this project:

1. What score(s) from the WIDA Screener Online should be used in identifying students for English language services?
2. What levels of performance (for each selected score) would indicate that a student does not need English language services?

The results of reach of these recommendations are described below.

Identification Criteria Scores

After initial ratings and discussion, the panel made their final recommendations as to which scores from the WIDA Screener Online should be included in the identification criteria. The results in Table 2 show the options discussed including the Overall Score only, the Overall Score along with the Composite Scores, the Overall Score along with the Domain Scores, or all scores (Overall, Composite, Domain). As shown in the table, the predominant recommendation was to use the Overall Score as well as the Domain Scores. Therefore, the groups were asked to conduct their work on the second meeting day considering this perspective.

Table 2. Identification Criteria Scores

Scores	Final Recommendation Voting
Overall Score Only	1 panelist recommended only using this overall score (compensatory model)
Overall Score + 1 or 2 Composite Scores	1 panelist recommended setting a minimum score requirement on the Oral Language composite in addition to the Overall score
Overall Score + Domain Scores	12 panelists recommended having a minimum score requirement for all domains in addition to the Overall score 1 panelist recommended having a minimum score requirement for just the writing domain in addition to the overall score
Overall Score + Composite Scores + Domain Scores	3 panelists recommended having minimum requirements for all scores

Performance Expectations

Table 3 includes the performance expectations from each grade level group after their initial discussion on Day 2 (independent recommendations made via online form), after their final discussion on Day 2 (independent recommendations made via online form), and then after the vertical articulation (as expressed by the grade-level representative). The values in red indicate a change in recommendation from the previous round. The values in the table indicate the combination of overall and domain scores that each panel recommended would indicate a student would not need English language scores. As an example, the kindergarten group believed that if a student achieved an overall score of 4.5 and domain scores of 4, this student would not need English language services. The full notes on each panel's rationale presented during the vertical articulation is included in Appendix A.

After the initial discussion, each group had a consensus on an overall score requirement and domain score requirements. Most groups (K, 2, 4-5, 6-8) set a specific expectation for the overall score that represented a general level of language proficiency with a slightly lower threshold for the domain scores to ensure a student did not have a specific area of need. The other groups (2-3, 9-12) set an initial expectation that was consistent between the overall score and domain scores. Given that the overall score represents an average of the domain scores, this perspective effectively focuses just on the domain performance (not the overall) as any student who achieves a score of 5 across all domains would automatically achieve an overall score of 5.

After reviewing the cross-state data and considering the perspective of the other grade level groups, three groups (K, 1, 9-12) maintained the same perspective whereas three groups (2-3, 4-5, 6-8) maintained their overall score expectation but were split on their domain score expectation.

Finally, after the vertical articulation, each group representative voiced their final opinion considering the discussions of their entire group as well as the perspectives of the other group representative. The two groups representing the youngest test takers (K, 1) recommended that any student who achieves an overall score of 4.5 and domain scores of 4 does not need English language services. Their key reasons included:

- Kindergarten: The group thought about the potential variation in pre-K experiences (some or none) and compared the performance expectations to what would be expected of students coming into kindergarten and what they would learn in the classroom. Believed the recommendations would achieve a balance between over- and under-identifying students.
- 1st Grade: The group believed that the Can-Do statements represent higher expectations, similar to what is expected in class. The group was also concerned that a domain score of 5 would over identify students and there would not be the resources available to meet the needs of all students, particularly those who need it most.

Three groups representing the middle grades (2-3, 4-5, 6-8) came to agreement that an overall score of 5 and domain scores of 4 would indicate that a student does not need EL services. Their key reasons included:

- 2nd – 3rd Grades: The group recognized the transition at this point where most students have been identified in earlier grades and the change in classroom focus from “learning to read” to “reading to learn”. Based on the Can-Do statements, the group felt an overall score of 5 and domain scores of 4 would indicate a student could function independently in the classroom.
- 4th – 5th Grades: The group considered how the expectations at each performance level aligned with the grade level expectations and academic standards. They believed that the score of 5 represented what



was needed to function independently in the classroom but that requiring all domain scores of 5 would be too high of an expectation.

- 6th – 8th Grades: The group believed that an overall score of 5 would represent a solid language proficiency but the expectation of a 4 represents the need for some support. Therefore, they thought a domain score of 4 represented a reasonable expectation (a “floor”) to ensure a minimal level of proficiency in each domain.

Finally, the high school representative recommended domain scores of 5 and recognized that this would also ensure a sufficient overall score and therefore did not recommend an overall score requirement. Their key reasons included:

- 9th – 12th Grades: The group recognized that this was the last opportunity students had to be identified for services while in school and therefore emphasized the importance of ensuring students had reasonable proficiency in each domain. Believed that a score of 5 represented that reasonable level of proficiency but that the overall score could mask a lower ability in one domain. Therefore, an overall score of 5 was not sufficient but domain scores would ensure this expectation is met.

Table 3. Performance Level Expectations

Grade Level	After Initial Discussion		After Final Discussion		After Vertical Articulation	
	Overall	Domain	Overall	Domain	Overall	Domain
K	4.5	4	4.5	4	4.5	4
1	4.5	4	4.5	4	4.5	4
2-3	5	5	5	4 (2 votes) 5 (1 vote)	5	4
4-5	5	4	5	4 (2 votes) 5 (1 vote)	5	4
6-8	5	4	5	4 (2 votes) 5 (1 vote)	5	4
9-12	5	5	5	5	<i>None</i>	5

Summary and Next Steps

As documented in this report, Washington educators engaged in a series of discussions and information sharing to formulate recommendations as to what level of performance on the WIDA Screener Online indicates that a student does not need EL services. Through iterative rounds of judgments, each grade-level panel established a general understanding, and the cross-group discussions yielded some consistency in recommendations (with a few slight differences). Overall, the panelists’ recommended expectations were slightly lower for kindergarten and 1st grade students and slightly higher for high school students.

These results and recommendations are presented to the Washington OSPI for their consideration and decision making.



Appendix A: Vertical Articulation

The notes below represent the final perspectives of each grade-level group as captured during the vertical articulation discussion and by the group facilitator.

Kindergarten

- The Can-do statements were really rigorous particularly when considering that students may or may not have any pre-K experience and lack cultural responsiveness. The Can-do statements appear more appropriate for classroom use and as their year-long goals towards 1st grade.
- The expectations within the Writing scoring rubrics and were very challenging and represented higher expectations than described in the Can-do statements.
- The domain scores of 5 would over-identify students given the potential lack of schooling prior to the screener (assume administration in the fall just before K). More concerned about under-identifying students given this will be the one and only opportunity for them to qualify for EL services. Believe that the combination of the overall score (5) and domain scores (4) would be a reasonable balance for over- and under-identifying.
- They also had to consider Transitional Kindergarteners, who are identified, based on district's own assessment, as needing additional support before entering kindergarten. These students are 4-year-olds so when considering their young age, many of them would highly likely qualify for EL services because they will be screened for measures designed for 5-year-olds.

1st Grade

- Can-do statements for Reading and Writing represented slightly high expectations
- Expect that students may do well with Oral Language but not Literacy
- Concerned about staffing to provide these services if OSPI over-identifies students for English language services. In turn, students that really need assistance may not receive it at the level needed. In turn, students who are identified but do not need services will miss out on grade-level services that they could greatly benefit from
- Believe that expecting domain scores of 5 across the four domains is more than would be expected from general education students at this grade and thus recommended a 4.

2nd – 3rd Grades

- Considered what level represents a student who can independently function in the classroom, ability to apply the skill vs learning the skill, ownership of the “can-do statements, and match between the Can-Do statements and the general education classroom activities.
- Most students are identified in at kindergarten and 1st grade
- Considered different domain score combinations that result in an overall score of 5
- Considered the transition from 1st to 2nd grade as students move from “learning to read” to “reading to learn”
- Reviewed the Screener scoring rubrics and felt students were able to meet the expectations
- Recognized different expectations and abilities across the Key Uses but thought 4 was a reasonable expectation when setting one expectation across all domains.



4th – 5th Grades

- Identified the performance indicators that align with the academic language demands of the common core standards
- Considered examples from the Can-Do statements at level 5-6, these levels represent what students will need to perform in the general classroom. The discussion was more focused on levels 4 and 5, and it was clear that the level 5 statements matched up with the academic language needed for full (unsupported) participation in general education classes
- One panelist shared SPED experience and discussed the possibility of not identifying students for SPED EL services but then getting identified for SPED (when they really need language services)
- Considered the risk of over identification vs students struggling in class if they are not identified. They wanted to err on the side of over identifying students.
- For the overall score, considered how partial points on the domain scores (which would be rounded down in the reporting of domain scores) will influence the overall score.
- Concerned about the Writing score being based on one task.

6th – 8th Grades

- Compared Can-Do statements between 4 and 5. The panel thought a score of 4 indicated the need for some support whereas a score of 5 showed the general consistency in language proficiency and what would be expected of all students coming into the grade level.
- Overall score of 5 would represent a solid language proficiency but the expectation of a 4 for each domain would represent a “floor” to ensure students have a reasonable proficiency in each domain
- Concerned with under identification and making sure the Screener caught all of the students who need the services. Did not consider the outside policy elements of staffing as this may vary from district to district

9th – 12th Grades

- Thought about civil rights and access to service and did not think about staffing, thought of ideal staffing and solution provision (i.e., did not worry about staffing if there is overidentification)
- Described the shift in expectations that occurs in grades 9-12 where students are asked to write to ensure demonstration of knowledge which complicates the thinking.
- Recognized that the expectations at Level 3 are not even close to grade level standard, this would only allow a student to participate in grade level with support and a 4 represents presumed knowledge which is never taught again. Concerned about linguistic task complexity required at HS are very high.
- Panel was concerned that if they set their expectation at 5 – for overall and domains - they are not accounting for individual learning differences. In other words, students may be strong communicators in some ways but not all (e.g., anxiety in a speaking session).
- Felt it was better to over-identify students and then those who do not need EL Services can test out in the spring rather than miss students.
- Overall want to ensure that a student does not have any 3’s at the domain level and not many 4s (i.e., want students to have mostly 5) but unsure of the overall score requirement that will ensure that expectation which is why they set the domain score expectation at 5.

